

2

Statistical Theory and Methods

Statistics allow workplace learning and performance (WLP) professionals to quantitatively describe and draw inferences about people, things, or events. In other words, statistics allow data to be organized and summarized and make it possible to draw generalizations and inferences. Statistics enable WLP professionals to document current levels of performance (individual, group, or organizational), measure the impact of their programs, and offer well-grounded feedback for change.

For many WLP professionals, the use of statistics is an onerous task—but it shouldn't be. Several software applications can do the number crunching for trainers; however, practitioners must understand how to use statistics. The selection and interpretation of statistics still rests in their hands. To use statistics properly, a statistics consumer needs to understand some essential concepts and principles. Although this chapter is designed to serve as a primer for statistics, occasionally calculations are used to help understanding.

WLP professionals must have a broad understanding of how data falls into distributions (for example, variance and normal distribution) and how data relates to other data (for example, correlation and regression). In addition, from an inferential standpoint, WLP professionals must understand concepts related to hypothesis testing, such as effect sizes and confidence intervals. It's easy to misuse or misinterpret statistics. Having a real understanding of statistics means that people can apply them correctly, represent findings accurately, and draw appropriate inferences.

Learning Objectives:

- Define and illustrate the three measures of central tendency.
- Define and compare the various types of frequency distributions.
- Express how measurement scales and statistical implications are used in the collection of measurement data.
- Explain how the measures of variance are used in statistics.
- Describe how distributions are used with standard scores.
- Identify the correct usage of correlation versus causation in data.
- List the five steps in the hypothesis-testing process.
- Demonstrate knowledge related to effect sizes and confidence intervals.
- Recognize the appropriate use of statistical information.

Measures of Central Tendency (Averages)

Besides using descriptive statistics for charts and graphs, trainers can perform several numeric calculations on them. The most common are called *measures of central tendency*, or *averages*, of which there are three: *mean*, *median*, and *mode*. Each type of average serves a unique purpose.

Mean

The *mean score* is considered the most robust, or least affected by the presence of extreme values (outliers), of the three types of central tendency measures, because each number in the data set has an impact on its (mean) value.

The mean is represented by the following formula:

Mean = Sum of all numbers divided by the number of values that make up the sum

The mean is a good measure of central tendency for roughly symmetric distributions but can be misleading in skewed, or nonsymmetric, distributions because it can be influenced a great deal by extreme scores. Therefore, other statistics, such as the median, may be more informative and appropriate for distributions that are often quite skewed, such as reaction time or family income.

Median

The *median* is the middle of a distribution arranged by magnitude: Half the scores are above the median, and half are below the median. The median is less sensitive to extreme scores than the mean, which makes it a better measure than the mean for highly skewed distributions. The median income is usually more informative than the mean income, for example.

The median in a distribution of odd- or even-numbered values, as noted in Table 1-1, is the calculated average of the two numbers of the high and low side of the *n*th number in the formula.

Which of these two measures—the sample mean or median—should trainers use? It depends. Usually medians are a better measure for skewed distributions than means, but this guideline must be tempered by common sense. In a dispute between the American Medical Association and the American Bar Association about the rising costs of malpractice insurance for doctors, the doctors used means to show a sharp rise in costs in the period 1980 to 1984, and the lawyers used medians to show that there was no rise at all (Schwarz 1998).

Mode

The *mode*, the most frequently occurring score in a distribution, is also used as a measure of central tendency. The advantage of the mode as a measure of central tendency is that its meaning is obvious. Further, it's the only measure of central tendency that can be used with nominal data.

The mode is greatly subject to sample fluctuations, so it's not recommended for use as the only measure of central tendency. A further disadvantage of the mode is that many distributions have more than one mode; these distributions are called *multimodal*.

The mean score is considered the most robust of the three types of central tendency because each number in the data set has an impact on the value of the mean. The median and the mode can be unaffected by individual numbers.

Table 2-1 shows a sample set of data collected by a trainer. These scores represent the pretest scores on a knowledge test a trainer administered before the start of a training program. This data is used to show the three types of measures of central tendency as well as variance.

Table 2-1. Pretest Scores (Arranged From High to Low)

Scores (x)	Mode (mo)	Median (mdn) ^a
29	The mode is the most frequently occurring number mo=37	mdn=(n+1) ÷2
35		mdn=(12+1) ÷2
37		mdn=6.5th number
37		mdn=54
46		
52		
56		
59		
61		
73		
77		
82		
$\Sigma x=644$		
The number of values, n, equals 12		
The mean = $\Sigma x \div n$		
The mean = $644 \div 12=53.67$		
Σx = sum of values in the list n = number of values in the list a Note that the formula in the example applies only to a case with an even number of values. For the odd number of values it is not necessary to average the two middle values to come up with the "virtual median." In the case of an odd number, the appropriate formula is $(n+1 \div 2) \div 2$.		

Frequency Distributions

A set of numbers may be summarized in two major ways: using pictures and using summary numbers. Each method has advantages and disadvantages, and the use of one method need not exclude the use of the other. This section describes drawing pictures of data called *frequency distributions*.

A frequency distribution can show the actual number of observations falling in each range or percentage of observations. With percentage of observations, the distribution is called a *relative frequency distribution*.

Some conditions that a frequency distribution might illustrate include dispersion, clusters, skewness, outliers, and normal distribution. The next section explores these concepts in more depth.

Definition of Terms

Confounding variable is an unknown or uncontrolled variable that produces an effect in an experimental setting. A confounding variable is an "independent variable" that the evaluator didn't somehow recognize or control. It becomes a variable that confounds the experiment.

Continuous variable is a variable whose quantification can be broken down into extremely small units (for example, time, speed, distance).

Control group is a group of participants in an experiment that's equal in all ways to the experimental group except that it didn't receive the experimental treatment.

Covariates are the multiple dependent variables in a study with multiple independent variables.

Dependent variable is frequently thought of as the "outcome," or treatment variable. The dependent variable's outcome depends on the independent variable and covariates.

Dichotomous variable is a variable that falls into one of two possible classifications (for example, gender [male or female]). An artificially dichotomous variable is imposed for classification purposes (for example, age classified as retired [>65] or not retired [<65]).

Discrete variable is a variable in which the units are in whole numbers, or "discrete" units (for example, number of children, number of defects).

Experimental group is the treatment group; those participants who receive the "treatment," for example, the training program.

Independent variable is the variable that influences the dependent variable. Age, seniority, gender, shift, level of education, and so on may all be factors (independent variables) that influence a person's performance (the dependent variable).

random error. The test can't consider biases resulting from nonrandom error (for example, a badly selected sample).

These are some key concepts about statistical significance:

- In statistical terms, *significant* does not necessarily mean important.
- Probability values should be read in reverse.
- Too many significance tests turn up some falsely significant relationships.
- It's important to check the sampling procedure to avoid bias.

Effect Sizes

Effect size is a way of quantifying the difference between two groups. For example, if one group (the treatment group) has had an experimental treatment and the other (the control group) has not, the effect size is a measure of the effectiveness between the two groups. Effect size uses standard deviation to contextualize the difference between the two groups.

Confidence Intervals

The *confidence interval* is the range where something is expected to be. Saying "expected" leaves open the possibility of being wrong. The degree of confidence measures the probability of that expectation to be true.

The degree of confidence is linked with the width of the confidence interval. It's easy to be very confident that something will be within a very wide range, and vice versa. Also, the amount of information (typically related to the sample size) has an influence on the degree of confidence and the width of the confidence interval. With more information, there can be more confidence that what's being measured will be within a given interval. Also, with more information and keeping a given degree of confidence, the interval can be narrowed.

For example, say a survey is conducted in Alexandria, Virginia. The question is "Do you prefer Coca-Cola or Pepsi?" Of the responses, 60 percent answer Coca-Cola, and 40 percent answer Pepsi. So the estimation is that, in this city, 60 percent prefer Coca-Cola. This doesn't mean that 60 percent of the population in this city prefers Coca-Cola—unless everyone in the population answered the survey. However, there's some "confidence" that the actual proportion of people choosing Coca-Cola will be within some interval around the 60 percent found in the sample. The amount of confidence depends on how wide the interval is. If the survey is based on a sample of 100 people, there can be 90 percent confidence that the actual proportion of those preferring Coca-Cola will be between 52 percent and 68 percent. Also, there can be 99 percent confidence that the actual proportion will be between 48 percent and 72 percent (for the same sample size, with more confidence and a wider interval). If the survey had been on a sample of 1,000 people instead of 100, there could be 90 percent confidence that the actual proportion is between 57.5 percent and 62.5 percent (compared with 52 percent and 68 percent for the same confidence with a sample of 100). Keep in mind that the larger the sample, the higher the degree of confidence.

Appropriate Use of Statistical Information and Data

As Mark Twain said, "Collecting data is like collecting garbage. Pretty soon, we have to do something with it." If not used properly, statistical information and evaluation data are useless. Improper use of evaluation data can lead to four major problems:

- Too many organizations don't use evaluation data at all. In these situations, data is collected, tabulated, catalogued, filed, and never used by any particular group other than the person who initially collected the data.
- Data is not provided to the appropriate groups. Different groups need different types of data and often in very different formats. Analyzing target audiences and determining the specific data needed for each group are important for communicating data.
- Data isn't used to drive improvement. Most evaluation data uncovers process improvement opportunities and identifies features that could be adjusted or changes that should be made to make the program more effective. If it's not part of the feedback cycle, evaluation falls short of what it's intended to do.
- Data is used for the wrong reasons—to take action against a person or a group or to withhold funds rather than improve processes. Sometimes the data is used in political ways to gain power or advantage over another person.

These problems represent dysfunctional activities that can destroy evaluation processes. They must be addressed if evaluation is to add value.

✓ Chapter 2 Knowledge Check

1. Which of the following best describes a situation where the mode < median < mean?
 - a. Negative skewness
 - b. Positive skewness
 - c. Outlier
 - d. Normal distribution
2. Which of the following best describes normal distribution?
 - a. An observation in a data set that's far removed in value from others in the data set
 - b. The symmetry in the distribution of the same data values
 - c. The way in which observations tend to pile up around the mean, also known as the bell-shaped curve
 - d. Variation in values that could be widely scattered or tightly clustered
3. Which of the following best describes dispersion?
 - a. An observation in a data set that's far removed in value from the others in the data set
 - b. The symmetry in the distribution of the same data values
 - c. The way in which observations tend to pile up around the mean, also known as the bell-shaped curve
 - d. Variation in values that could be widely scattered or tightly clustered
4. Which of the following best describes an outlier?
 - a. An observation in a data set that's far removed in value from the others in the data set
 - b. The symmetry in the distribution of the same data values
 - c. The way in which observations tend to pile up around the mean, also known as the bell-shaped curve
 - d. Variation in values that could be widely scattered or tightly clustered
5. Which of the following types of data make it possible to rank order items measured in terms of which has less or more of the quality represented?
 - a. Nominal
 - b. Ordinal
 - c. Interval
 - d. Ratio

6. Which of the following types of data include the feature of identifying an absolute zero point?
 - a. Nominal
 - b. Ordinal
 - c. Interval
 - d. Ratio
7. Variance is defined as how spread out a distribution of data points is, whereas the standard deviation is the measure of how spread out the data points are when the mean is used to calculate central tendency.
 - a. True
 - b. False
8. The reason that practitioners convert raw scores to standard scores includes which of the following?
 - a. To indicate the number of correct answers to allow scores to be compared
 - b. To reflect where they fall with respect to the mean to allow scores to be compared and interpreted
 - c. To understand the cause-and-effect connections between variables
 - d. Because they are always expressed as a number between -1.00 and $+1.00$
9. An example of a relational study statistic that measures the relationship between two or more variables includes
 - a. Correlation coefficient
 - b. Cause-and-effect connection
 - c. Normal distribution
 - d. Skewness
10. The primary goal of hypothesis testing is to test a hypothesis and then accept or reject the hypothesis based on the findings.
 - a. True
 - b. False